Kidney-specific methylation patterns correlate with kidney function and are lost upon kidney disease progression

Naor Sagy[1], Noa Meyrom[1], Pazit Beckerman[2], Oren Pleniceanu[2] and Daniel Z Bar[1,3]

1 - Department of Oral Biology, The Goldschleger School of Dental Medicine, The Faculty of Medicine, Tel Aviv University 69978, ISRAEL
2 - Kidney Research Lab, The Institute of Nephrology and Hypertension, Sheba Medical Center, Tel-Hashomer and The Faculty of Medicine, Tel-Aviv University
3 - The AI and Data Science Center (TAD), Tel Aviv University 69978, ISRAEL

Abstract
Chronological and biological age correlate with DNA methylation levels at specific sites in the genome. Linear combinations of multiple methylation sites, termed epigenetic clocks, can inform us of the chronological age and predict multiple health-related outcomes. However, why some sites correlate with lifespan, healthspan, or specific medical conditions remains poorly understood. Kidney fibrosis is the common pathway for Chronic Kidney Disease, which affects 10% of Europe and USA population. Here we identify epigenetic clocks and methylation sites that correlate with kidney function. Moreover, we identify methylation sites that have a unique methylation signature in the kidney. Methylation levels in the majority of these sites correlates with kidney state and function. When kidney function deteriorates, all of these sites regress towards the common methylation pattern observed in other tissues. Interestingly, while the majority of sites are less methylated in the kidney and become more methylated with loss of function, a fraction of the sites are highly methylated in the kidney and become less methylated when kidney function declines. These methylation sites are enriched for specific transcription-factor binding sites. In a large subset of sites, changes in methylation pattern are accompanied by changes in gene expression in kidneys of chronic kidney disease patients. These results support the information theory of aging, and the hypothesis that the unique tissue identity, as captured by methylation patterns, is lost as tissue function declines. However, this information loss is not random, but guided towards a baseline that is dependent on the genomic loci.

Introduction

Chronic kidney disease (CKD)

The human kidney plays several essential roles, including the excretion of toxic waste products and maintenance of blood pressure and pH. The average kidney contains ~1 million nephrons, which are its main functional units (Fairweather et al. 2020). CKD is defined as loss of renal function, marked by decreased glomerular filtration rate (GFR)<60 mL/min/1.73m², and increase in kidney damage markers (e.g., albuminuria), or both, for ≥3 months (Anon 2013). Renal function can be assessed using serum creatinine (sCr) levels, used to compute the estimated Glomerular Filtration Rate (eGFR) via one of several formulae (Musso et al. 2016). Unfortunately, sCr-based assessments are highly inaccurate due to several reasons: (1) sCr is highly insensitive for diagnosing early CKD, as abnormalities in its values occur only when ~40% of the renal parenchyma is damaged (Steubl et al. 2016). Indeed, many individuals with normal sCr actually have kidney damage upon histological examination including interstitial fibrosis (IF), tubular atrophy (TA) and glomerulosclerosis (GS) - all indicative of impaired renal function (Mancilla et al. 2008; Okamoto et al. 2010; El Agha et al. 2017). CKD is a relentless disorder, steadily deteriorating from its early stages (1-2) in which the GFR is normal or high to stage 5, or end-stage kidney disease (ESKD), defined as a GFR<15 ml/min/1.73m2, when patients require kidney replacement therapy (KRT): either dialysis or transplantation (Anon 2013). With a prevalence of up to 17.3% (Brück et al. 2016) requiring expensive treatments, CKD is a global epidemic, accounting for over

1.2 million deaths a year (GBD 2015 Mortality and Causes of Death Collaborators 2016). Moreover, its true prevalence is hard to determine, since its early stages often go undetected. CKD rates are expected to keep rising due to the aging of the population and growing prevalence of its 2 main risk factors: diabetes mellitus (DM) and hypertension (HTN) (GBD 2015 Mortality and Causes of Death Collaborators 2016). The irreversible nature and lack of specific treatments for CKD underscore the importance of its early diagnosis, while renal function is preserved and active interventions (e.g., glycemic control in DM) may slow its progression, which will improve quality of life, reduce mortality, and reduce the costs for healthcare systems (Levin & Stevens 2011).

DNA methylation correlates with lifespan, healthspan and CKD

DNA methylation is an epigenetic modification, and the most studied one in relation to diseases, healthspan and lifespan. Seminal works by Horvath and Hannum described 'epigenetic clocks', a linear combination of the methylation levels of several tens to hundreds of genomic sites, which accurately estimate the biological age of most human tissues, including the kidney (Lu et al. 2019; Horvath 2013; Hannum et al. 2013). These clocks have been expanded to predict "biological age", capture multiple aspects of healthspan and predict all-cause mortality (Lu et al. 2019). Proving the relevance of epigenetic signatures to renal function, the Susztak group profiled genome-wide methylation of tubular cells in CKD and normal kidneys and found that a 65-probe signature correlates with kidney structural damage, as seen in biopsy, and a 471-probe signature predicts renal functional decline together with clinical parameters (Gluck et al. 2019). Likewise, it has been recently shown that the progression of CKD is tightly linked to methylation-induced changes in tubular cell function, including senescence (Al-Dabet et al. 2022). Remarkably, renal methylation signatures also accurately reflect the presence of kidney cancer, with the methylome of RCC cells exhibiting a significantly older biological age, a phenomena seen in a wide range of cancers. Moreover, it has been shown that specific DNA methylation signatures predict various clinical outcomes, and correlate with renal function (Lu et al. 2019; Dritsoula et al. 2021; Rysz et al. 2022; Schlosser et al. 2021).

Changes in DNA methylation alter transcription factor binding

DNA methylation often correlates with transcription levels. Epigenome-wide association studies (EWAS) demonstrated that methylation at specific CpGs sites are associated with kidney disease (Wing et al. 2014; Chu et al. 2017; Ko et al. 2013; Sapienza et al. 2011; Smyth et al. 2014; Rakyan et al. 2011). Methylation and demethylation can have direct effects on expression levels *in vivo* (Kangaspeska et al. 2008; Métivier et al. 2008) thus suggesting that some of the correlations might have a causal component to them. Methylation signatures can alter some TFs binding (Bonder et al. 2017; Héberlé & Bardet 2019), which can mechanistically explain some of this effect.

Here, we identify epigenetic clocks and CpG sites that correlate with the functional state of the kidney, as captured by IF and eGFR. A subset of these CpG sites show a kidney specific methylation pattern. Moveover, the majority of sites showing a kidney specific methylation pattern correlate with IF. Multiple genes associated with these sites are changed in CKD. Interestingly, all these sites lose their unique methylation pattern in CKD, while trending towards the common form.

Results

Specific epigenetics clocks correlate with eGFR and IF

CKD progression is accompanied by a decrease in eGFR and an increase in IF (Mancilla et al. 2008; Okamoto et al. 2010; El Agha et al. 2017). Several studies have analyzed the methylation of DNA in kidney tissue and identified CpGs whose methylation levels correlate with these markers of kidney disease progression (Schlosser et al. 2021; Ko et al. 2013; Bontha et al. 2017). To test whether eGFR decrease and IF increase are captured by existing epigenetic DNA methylation

clocks, we analyzed 85 publicly available kidney methylation arrays (Ko et al. 2013; Bontha et al. 2017). Multiple clocks showed a significant correlation with eGFR and IF (**Supp. Fig. 1**; **Supp. Table 1**). We noticed substantial differences in specific clocks performed between males and females, thus we performed these analyses independently (**Fig. 1,2**; **Supp. Tables 2-3**). Of all clocks tested, DNAmB2M performed best in males, giving a correlation of -0.55 ($p=4\times10^{-4}$) to eGFR and 0.75 ($p=7\times10^{-8}$) to IF. By contrast, in females, DNAmPhenoAge gave a correlation of -0.39 (p=0.017) to eGFR and DNAmPAI1 of -0.39 (p=0.015) to IF. Age and BMI, two potential confounders, were poorly correlated with these clocks (**Supp. Tables 2-3**). We concluded that some aspects of the physiological state of the kidney are reflected in the methylome and captured by epigenetic clocks not specifically developed for this purpose.

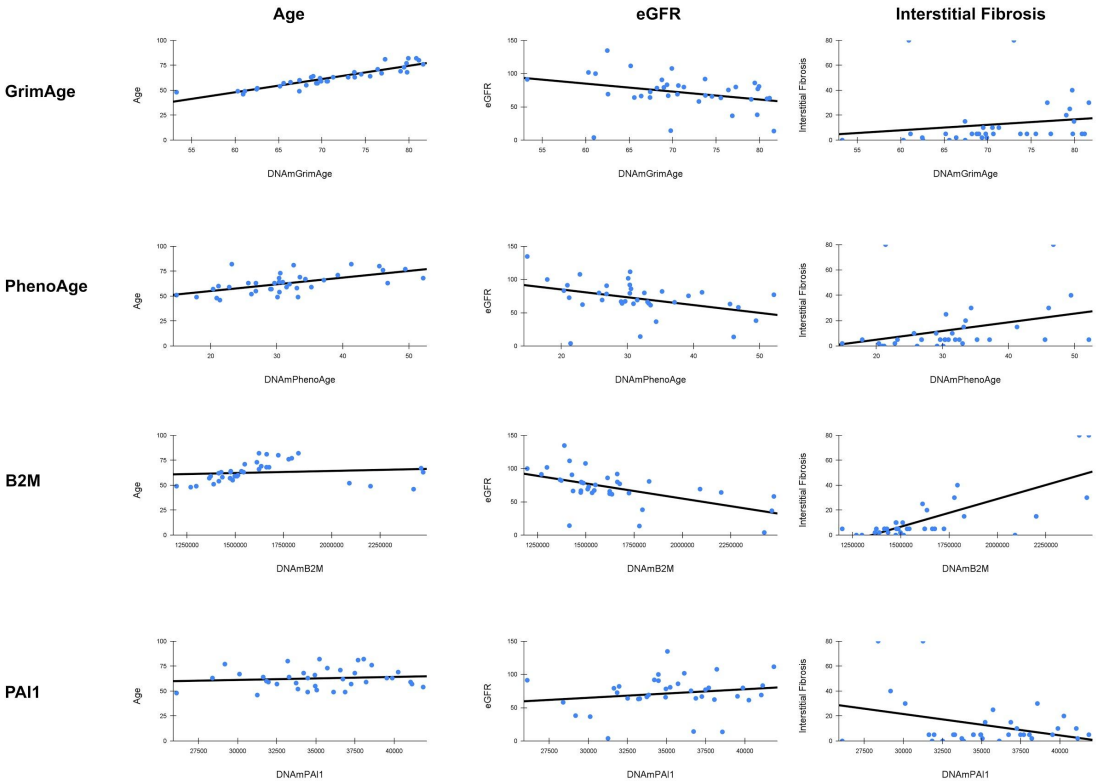


Figure 1: Methylation clocks correlate with kidney state in males. Correlation of age, eGFR and IF with four epigenetic clocks: GrimAge, PhenoAge, B2M and PAI1.

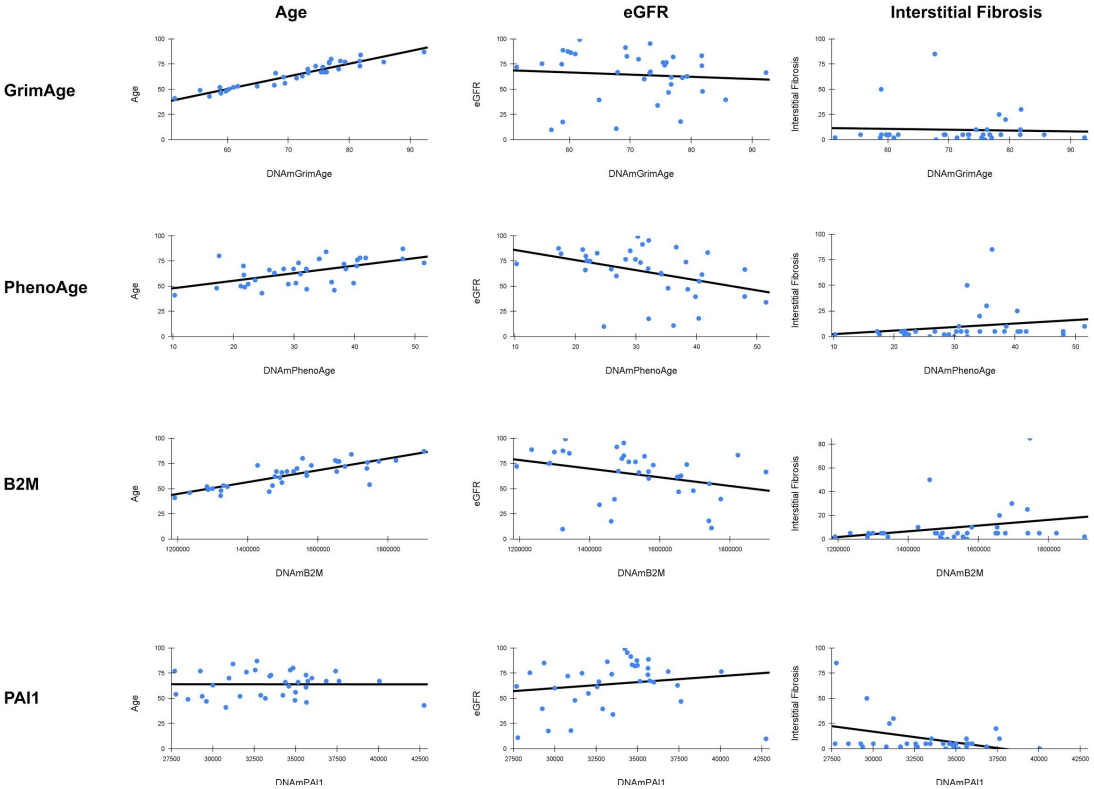


Figure 2: Methylation clocks correlate with kidney state in females. Correlation of age, eGFR and IF with four epigenetic clocks: GrimAge, PhenoAge, B2M and PAI1.

Multiple CpG sites show strong correlation to eGFR and IF

Alternative methylation sites may provide an even better estimate of tissue state. Despite the small number of arrays, 1975 sites showed a statistically significant correction (after Bonferroni correction) for eGFR, with correlation values up to r=0.74 (**Fig. 3A,B**, **Supp. Table 4**). We tested if these results can be explained by simple global loss or gain of methylation. Of the 1975 sites, 42% positively and 58% negatively correlated with eGFR, excluding global unidirectional changes as drivers of these methylation patterns. For IF, 20,986 sites passed Bonferroni correction, with r as high as 0.89 (**Fig. 3C,D**, **Supp. Table 5**). Of these, 67% showed a positive correlation and 33% a negative correlation. To test for the possibility that age, sex, race or BMI act as confounders, we measured their correlation eGFR and IF in this dataset. Additionally, we examined the correlation of top CpG sites with age and BMI in this and the independent NGDC-CNCB (Xiong et al. 2021) dataset. In all cases, correlation to age, sex and BMI was weak or negligible, excluding them as confounders (**Supp. Table 6**). We took a closer look at cg10832035, the site with the highest correlation to eGFR (r=-0.74; $p<10^{-13}$). As expected, it also showed a strong correlation to IF (r=0.77, $p<10^{-14}$). It was located on a CpG island inside the non-coding gene RP5-1086L22.1. The adjacent methylation sites within the island, but not outside the island, exhibited a similar methylation pattern and correlation to IF (r=0.7; $p<10^{-10}$) and eGFR (r=-0.64; $p<10^{-8}$), excluding possible probe issues. Interestingly, this site was almost completely methylated in all tissues except the kidney (**Fig. 3E**). In the kidney, methylation levels averaged around 0.35, and tended to increase with IF and decline in eGFR.

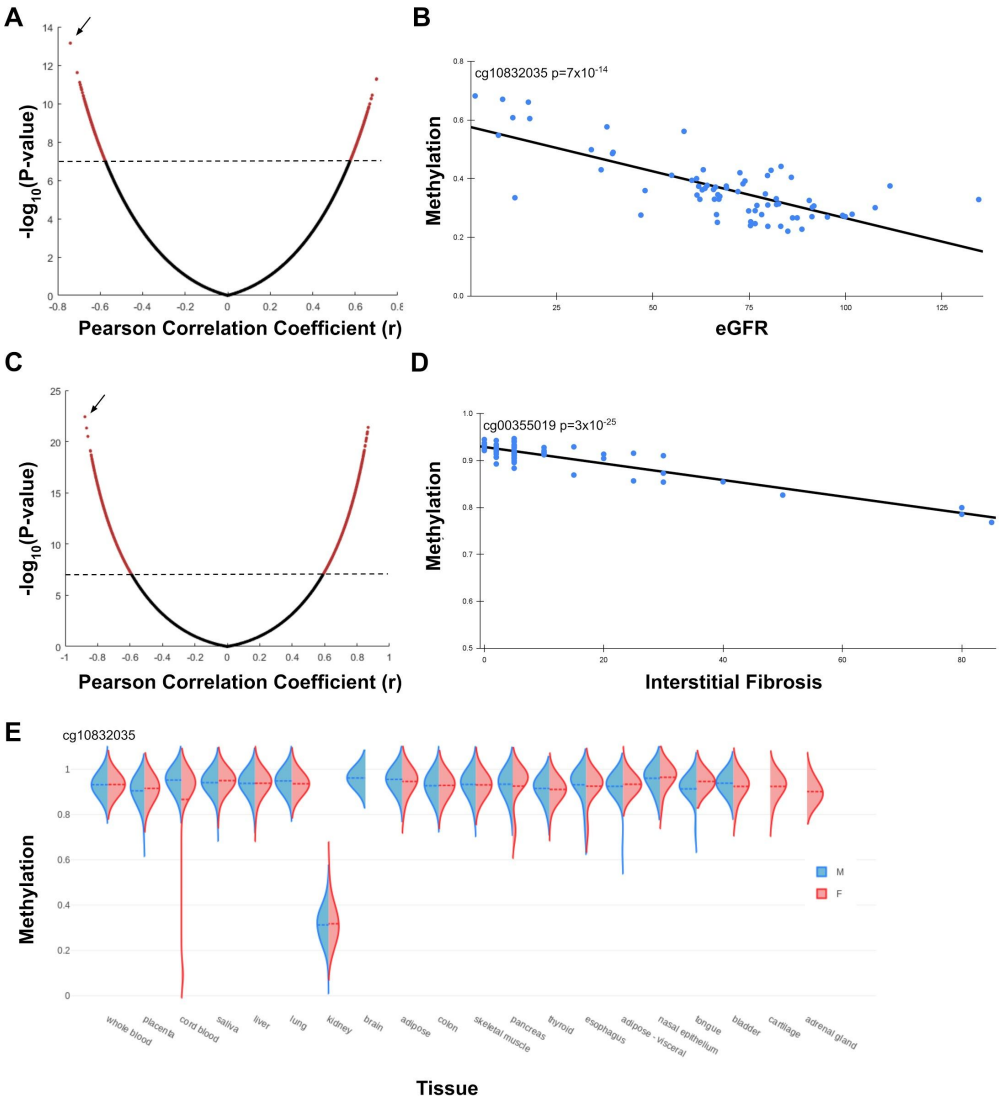


Figure 3: Individual CpG sites correlate with kidney state. A. Volcano plot of all tested CpG sites and their correlation to eGFR. Dots above the dashed line (red) pass Bonferroni correction. **B**. Individual methylation and eGFR levels for cg10832035, the site with highest correlation to eGFR (arrow in A). **C**. Volcano plot of all tested CpG sites and their correlation to IF. **D**. Individual methylation and eGFR levels for cg00355019, the site with highest correlation to IF (arrow in C). **E.** cg10832035 methylation patterns in a subset of tissues from the NGDC-CNCB dataset.

CKD-correlated CpG sites are enriched for TF binding sites

To test for potential biological significance, sequences surrounding these CKD-correlated sites were used for motif enrichment analysis. Indeed, these sequences were enriched for transcription factor (TF) binding sites. Specifically, sites that positively correlated with IF showed enrichment for the HNF1 family, while negatively correlated sites were enriched for Jun/FOS (**Fig. 4**).

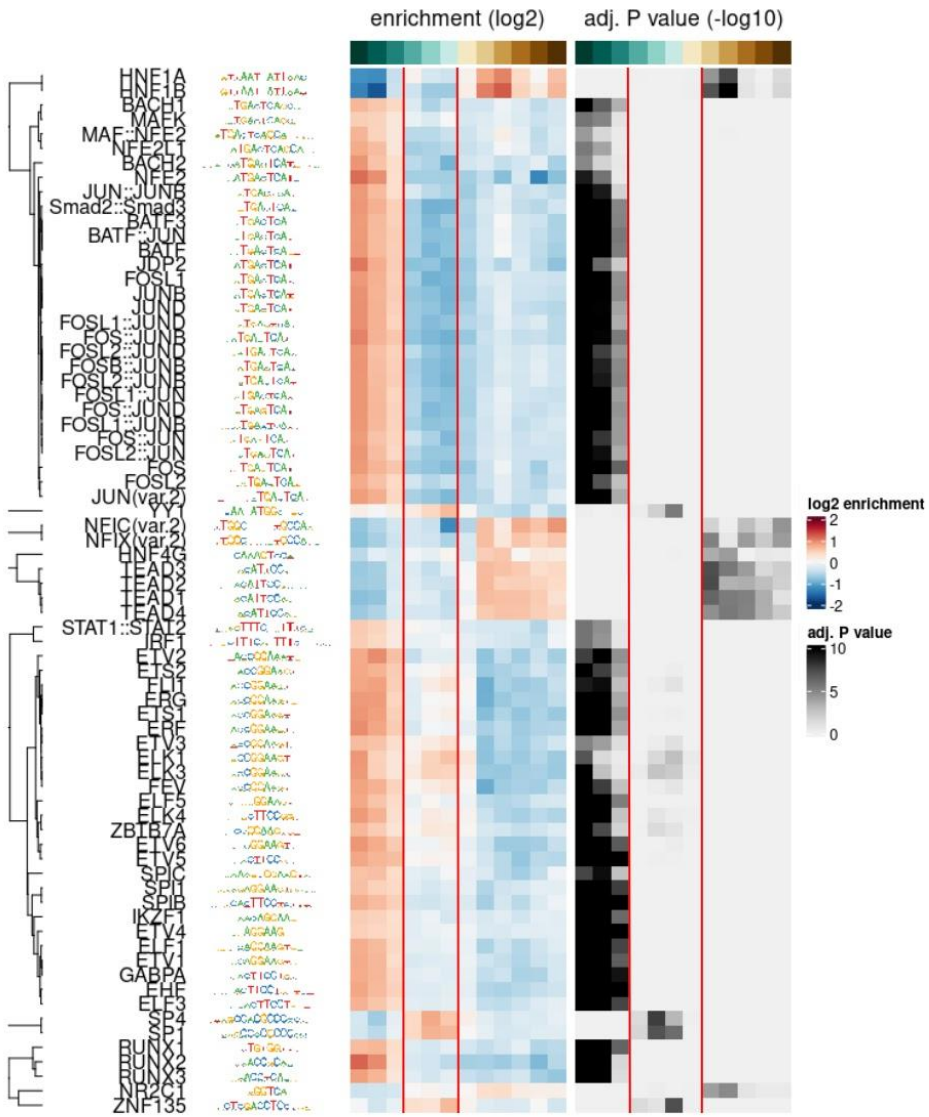


Figure 4: Methylation sites that correlate to IF are enriched for specific TF binding sites. TF enrichment and p-values of sequences around CpG sites that positively correlate with IF (left), do not correlate with IF (middle; between the red lines) and negatively correlate to IF (right).

Kidney-specific methylation patterns are lost with kidney disease progression

Tissue-specific differentially methylated regions (tsDMRs) can have tissue-specific roles (Irizarry et al. 2009; Davies et al. 2012). After noticing the unique methylation pattern of cg10832035, we sought to test if this is a general phenomenon. Using the NGDC-CNCB (Xiong et al. 2021) dataset, we identified 427 methylation sites that exhibit a unique methylation pattern in the kidney (**Supp. Table 7**). Of these sites, 68% were in the top 10% ($r>0.47$; $p<4\times10^{-5}$) of sites correlated with IF. Removing CpG sites with significant missing data increased this percentage to 72% (208 of 289; expected: 29; $p=10^{-322}$; **Supp. Table 7**). Thus, the overwhelming majority of kidney unique methylation sites change their methylation pattern with functional decline.

Function decline is accompanied by a drift towards the common methylation pattern

Cg10832035 is partially methylated in the kidney, but almost completely methylated in every other tissue examined (**Fig. 3E**). Indeed, we found this to be the common case. Of the 289 sites analyzed, 253 (88%) had a lower methylation level in the kidney as compared to the rest of the tissues, while only 36 (12%) had a higher methylation level. Upon kidney functional decline, cg10832035 becomes more methylated, trending towards the common methylation state of all other tissues examined. More generally, this raises three distinct possibilities, that upon functional decline (i) methylation levels increase (ii) methylation levels trend towards the common form (iii) methylation may increase or decrease independently of the common form. To distinguish between these three possibilities, we examined the levels of methylation in kidneys with different levels of

fibrosis. For all 289 sites, methylation level trended toward the common form in the body ($p=10^{-87}$), increasing in the 253 sites that are less methylated in the kidney compared to other tissues (**Fig. 5A**) and decreasing in the 36 sites that are more methylated in the kidney (**Fig. 5B**). Moreover, there was a positive correlation between the "uniqueness" of each site (as measured by the distance from all other tissues) and the correlation of the site with IF score (**Fig. 5C**; r=0.95 $p<10^{-148}$). We concluded that kidney functional decline was associated with an erosion of the unique methylation pattern of the kidney.

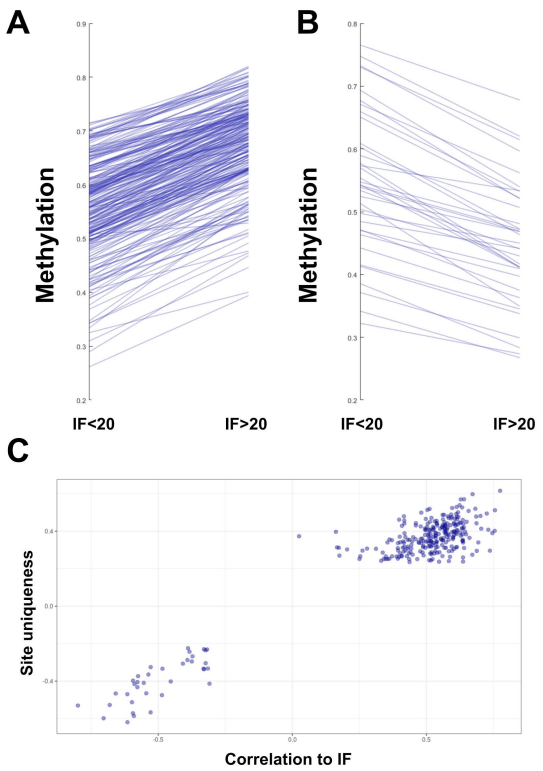


Figure 5: Methylation levels regress to the common baseline upon CKD. A. Methylation levels for the 253 sites undermethylated in the kidney, in individuals with IF<20 and >20. **B.** Methylation levels for the 36 sites overmethylated in the kidney, in individuals with IF<20 and >20. **C.** Site uniqueness, as measured by the distance from other tissues, and the corresponding correlation to IF.

Characterization of kidney-specific methylation sites

Methylation patterns may be influenced by protein-DNA interaction, which in turn are guided by the primary DNA sequence. To identify motifs that may help establish or maintain these unique methylation patterns, we extracted the sequences surrounding the 253 and 36 identified sites, and looked for motif enrichment. *De novo* motif discovery, using random CpG sites as background, identified multiple sequences showing strong enrichment (**Supp. Fig. 2**). Not surprisingly, these included significant enrichment for motifs identified in CpGs that correlated with CKD progression. To identify sequences that potentially contribute to the unique methylation pattern, while removing general CKD background, we repeated this analysis using as background CpG sites that correlate to IF, but do not have a unique methylation pattern. Again, multiple motifs were identified, including HNF1 and HIF2a (**Supp Fig. 3**). However, the 253 sites identified significantly differ from both backgrounds. As most sites are almost fully methylated in all but one tissue, they may differ in GC content from the background. Indeed, the average GC percentage of both backgrounds, as well as the methylated-in-kidney sites, was ~55%. By contrast, the average GC content around the 253 sites (predominantly methylated in the body, with an exception in the kidney) was 45%. Thus, we compared these sites showing a unique methylation pattern, but in other tissues. Again, we saw a significant enrichment for specific TFs, including HNF1 (**Supp. Fig. 4**).

Expression change of genes associated with kidney-specific methylation patterns

Changes in methylation levels are often associated with changes in gene expression. Depending on the genomic context, these changes can be causal or non-causal, and can involve an upregulation or downregulation of the associated gene (Dhar et al. 2021; Moore et al. 2013). We tested if changes to kidney-specific methylation patterns are associated with gene expression changes. To achieve this aim, we identified genes associated with the identified CpGs (**Supp. Table 7**), and tested for changes in mRNA levels in kidney biopsies of CKD patients and controls (Nakagawa et al. 2015). Of the 250 genes tested, 210 (84%) showed a statistically significant ($p<0.05$) difference between CKD and control (**Supp. Table 8**). Of these ~90% were upregulated in CKD (**Fig. 6**; **Supp. Table 8**). We concluded that methylation in these sites correlates with gene expression, likely by reflecting local chromatin structure.

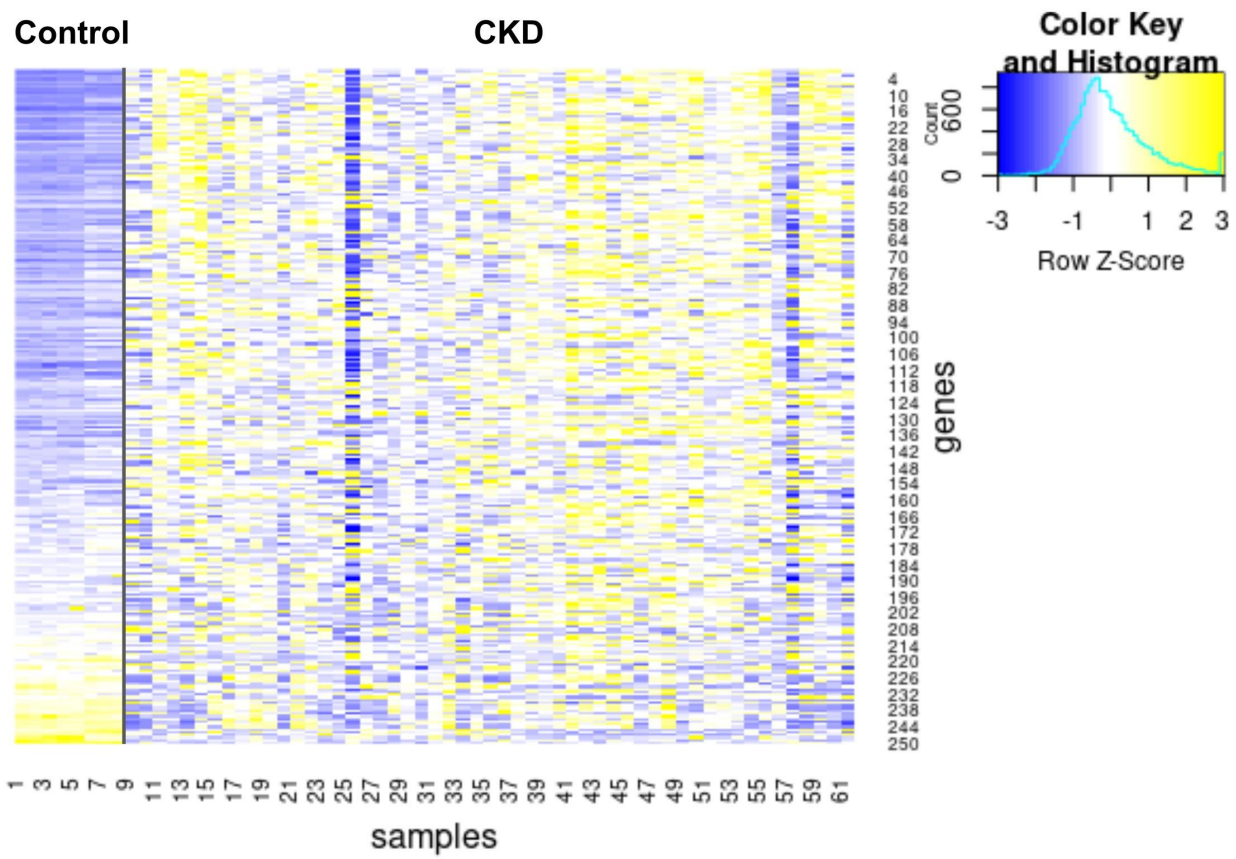


Figure 6: Genes associated with uniquely methylated sites change in expression in CKD. Expression levels of all (250) genes associated with uniquely methylated sites. Gene expression from the Nakagawa CKD Kidney dataset, with the discovery and validation cohorts united.

Methylation sites negatively correlated with IF are linked to oncogenes and developmental genes

We clustered the 36 methylation sites negatively correlated with IF into 15 genomic regions which mapped to 13 genes (**Supp. Table 9**). Of these 13 genes, seven are known as renal cancer prognostic genes (overexpression), five take role in kidney differentiation and development, two of which overlap with cancer prognostic genes, and one functions as a tumor suppressor. Kidney fibrosis is known to correlate with CKD and subsequently with renal cancer (Lowrance et al. 2014; Christensson et al. 2013). SPAG5, MCF2L, CCDC64, PDLIM4, EMX1, P4HA2, VIM and GRAMD1B methylation declined for IF>20, which might indicate oncogenesis in the fibrotic tissue. Interestingly, CKD is a well-established risk factor for renal cancer (Lowrance et al. 2014; Christensson et al. 2013). Although kidney inflammation (Silverstein 2009) and relative immunodeficiency (Giacchino et al. 1982) have been suggested to play a role in CKD-related carcinogenesis, the exact molecular link has yet to be defined. Notably, although CKD results in a generalized inflammatory state (Shlipak et al. 2003), the risk of cancer in CKD has been shown to be limited to the kidney and not other organs (Silverstein 2009), implying that it likely arises due to kidney-specific changes. The described methylation changes, which are highly kidney-specific, may thus account for at least some of the enhanced cancer risk among CKD patients.

Epigenetic information loss correlates with IF and eGFR

The discovery that epigenetic information is lost in a specific subset of methylation sites opens us the question: can a pathological state be predicted from methylation data alone, without any other prior knowledge. To answer this question, we implemented a naive information loss score. For each of the 289 unique methylation sites, discovered in the NGDC-CNCB dataset, the kidney median and standard deviation were calculated. Next, for each individual, the number of sites that

deviate from the tissue median towards the levels in other tissues, by at least two standard deviations, was calculated (**Fig. 7**). This score naively represents the loss of epigenetic information in a specific kidney sample. The information loss score showed a r=0.82 correlation with IF ($p<10^{-18}$) and r=-0.54 with eGFR ($p<10^{-6}$). Thus, a pathological condition can be probabilistically observed merely by noting the loss of epigenetic information, without any prior knowledge of the disease to affected tissue.

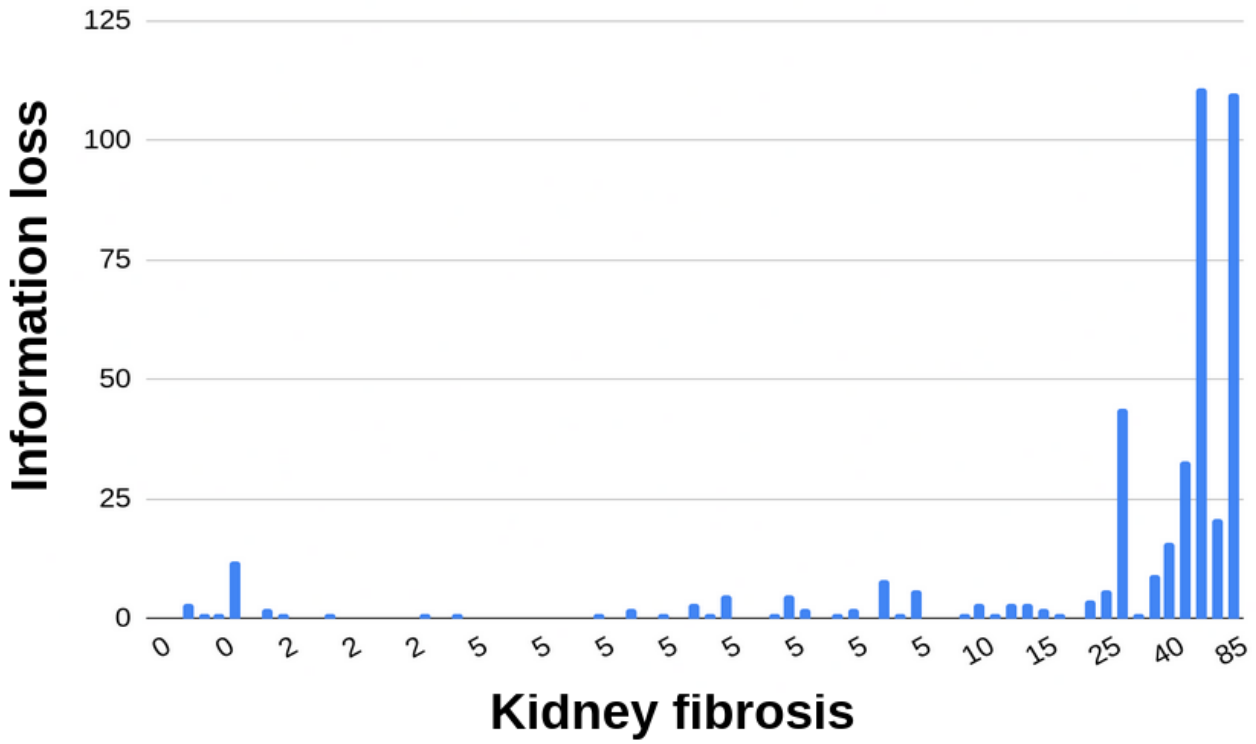


Figure 7: Epigenetic information loss correlates with IF. Epigenetic information loss score for each individual of the GSE50874 dataset (y-axis) along with the corresponding IF score (x-axis).

Discussion

Kidney state is captured by specific CpG methylations

Analysis of the kidney methylome revealed that multiple kidney pathologies are captured by distinct epigenetic clocks (**Fig. 1,2**; **Supp. Fig. 1**). DNAmB2M and DNAmPAI1 had a particularly good correlation with IF in males (r=0.75 and 0.58 respectively). DNAmB2M clock was built based on Beta-2 Microglobulin levels (Lu et al. 2019). Increased B2M levels in the blood, and by contrast low levels in the urine, are indicative of glomerular dysfunction (Argyropoulos et al. 2017; Sedighi et al. 2015; X et al. 2014), suggesting a possible explanation. Of note, DNAmB2M and DNAmPAI1 showed relatively low correlation amongst themselves (r=-0.30 in males and -0.14 in females), indicating that an improved estimator can be generated to facilitate more accurate estimation of kidney function across both sexes. Indeed, a global analysis identified multiple methylation sites, showing strong correlations for both sexes. A future increase in the number of samples will enable the construction of new epigenetic clocks, capturing kidney function decline with greater precision.

Next, we identified individual CpG sites that correlated with IF or eGFR. As expected, IF and eGFR associated CpGs showed a significant overlap. On average, IF resulted in stronger correlations than eGFR both for methylation clocks and specific CpGs. This is also expected, as eGFR is an inherently noisy measurement that does not capture the kidney state as accurately as IF (Levey et al. 2009). As expected, these results greatly overlapped with Bontha et al. 2017. By contrast, only

a small overlap of 21 of 69 CpG sites was observed when compared to blood methylation markers of CKD (Meta-analyses identify DNA methylation associated with kidney function and damage). We concluded that the physiological state of the kidney is reflected by the methylation levels at multiple sites.

Tissue-specific methylation patterns are lost with functional decline

Of all sites passing our rigorous threshold for tissue-specificity, 72% have a methylation pattern that correlates with IF (**Supp. Table 7**). This is likely an underestimation, as the arbitrary cutoff of 10% ($p<4\times10^{-5}$) is guided by the relatively small dataset used to identify correlated CpGs. The loss of unique epigenetic signature with functional decline appears to be nearly universal in the kidney. Loss of epigenetic information has been suggested to accompany and potentially drive cellular aging (Lu et al. 2020; Imai & Kitano 1998). These findings support this hypothesis, and point that some of the loss may correlate better to functional decline or “biological age” than to chronological age. Moreover, tissue-specific methylation sites appear to be particularly vulnerable to such loss, as most uniquely methylated sites were associated with functional decline, compared to much smaller numbers among all sites. This exceptionally high occurrence leads us to speculate maintenance of these unique sites may serve as a hallmark of functional tissue. Future studies will test if this phenomenon also applies to other tissues, as well as functional decline that does not involve fibrosis.

Uniquely methylated sites revert to the average

Interestingly, in every single uniquely methylated site tested, the functional decline was accompanied by shift in methylation levels towards the levels common in other tissues. The most plausible explanation for this would be that a default pattern exists in these sites, and a deviation from this default requires active maintenance. Primary DNA sequence is a key driver of local chromatin architecture. The unique methylation patterns could be established and maintained by specific DNA motifs. We tested for DNA motifs that could explain these patterns. While no single motif could explain this unique pattern, we have identified several candidates, enriched in these sites. *De novo* motif identification discovered not only motifs depleted near unrelated CpG sites, but also motifs present in uniquely methylated sites correlating to IF, but depleted near non-uniquely methylated sites that do correlate with IF.

While this work focused only on kidney-specific methylation patterns, we did observe multiple sites that were shared with one or a few other tissues and were excluded from our lists. It is likely that these findings can be greatly expanded, both by testing sites that show a distinct methylation pattern in other tissues, and by analyzing sites that show a distinct methylation pattern in a subset of tissues.

Gene expression correlates with changes in methylation

DNA methylation can reflect aspects of local chromatin structure, thus only correlating to gene expression. Alternatively, DNA methylation alters the binding for some TF (Héberlé & Bardet 2019), affecting gene expression. We identified multiple genes, associated with CKD-correlated CpG sites, that change in expression in CKD (**Fig. 6**; **Supp. Table 8**). CKD-correlated CpG sites are enriched for JunD TF binding sites. DNA methylation inhibits JunD binding (Mann et al. 2013; Spruijt et al. 2013), thus providing a potential causal mechanism by which methylation at these sites regulates gene expression.

Materials and methods

Epigenetics clocks and CpG site correlations

Epigenetic estimators were calculated using the DNA Methylation Age Calculator (Horvath 2013) for the 85 Illumina HumanMethylation450 BeadChip array samples for which full pathological data was available (GEO GSE50874). Samples that generated the warnings "meanMethBySample <0.25" were removed and the data was re-analysed (**Supp. Table 1**), resulting in a very mild increase in most correlations. Next, samples separated into males and females and analyzed again (**Supp. Table 2-3**). All CpG sites were scored by Pearson correlation (r) to eGFR and IF. For each r, $t=r\sqrt{(n-2)} / \sqrt{(1-r^2)}$ was calculated, along with the corresponding p-value. Age and sex were were excluded as confounders by testing for association with eGFR and IF both the GSE50874 and the NGDC-CNCB dataset.

Identification of uniquely methylated sites

The National Genomics Data Center, China National Center for Bioinformation (NGDC-CNCB; Xiong et al. 2021) dataset contains tissue-level methylation data across multiple partially annotated individuals. It was used to infer tissue-specific methylation levels. Uniquely methylated sites were defined as such that the average methylation level in kidney is lower by at least 0.2 than the 5% quantile level of all other samples in the NGDC-CNCB dataset, or higher by at least 0.2 than the 95% quantile level of all other samples. This distinguished methylation sites that out of the 28 tissues in the dataset, have a unique signature in the kidney, but did not exclude samples where a very small minority of samples leaks throughout the entire range. Next, CpG sites with less than 2000 values were removed. These sites were identified using custom R and Python scripts.

Motif Enrichment

Motif enrichment analysis for CKD-correlated CpG sites was done using MonaLisa (Dania Machlab, Lukas Burger, Charlotte Soneson, Michael Stadler 2021), using 200bp of sequence from each side of each CpG site. As a control, CpG sites showing no correlation were inserted between the positively and negatively correlated sites. *De Novo* identification of motifs near uniquely methylated sites was done using Homer (Heinz et al. 2010) on 500bp of sequence from each side of each CpG site. These were compared to randomly selected CpG sites, or IF correlated randomly selected CpG sites.

Gene expression analysis

Genes associated with all uniquely methylated sites CpG sites were extracted from the HumanMethylation450 v1.2 Manifest File (Illumina, USA). Gene expression levels in CKD and control kidney biopsies from the Nakagawa CKD Kidney dataset (Nakagawa et al. 2015) were extracted using NephroSeq.

Author contribution

DZB and NS designed the experiments. NS performed the experiments with inputs form NM, PB, OP and DZB. NM performed the monaLisa analysis. DZB wrote the manuscript with inputs form NS, PB and OP.

Funding

The work was supported by the Israeli Science Foundation (grants 654/20 and 632/20 to DZB).

Conflict of interest

The authors declare no conflict of interest.

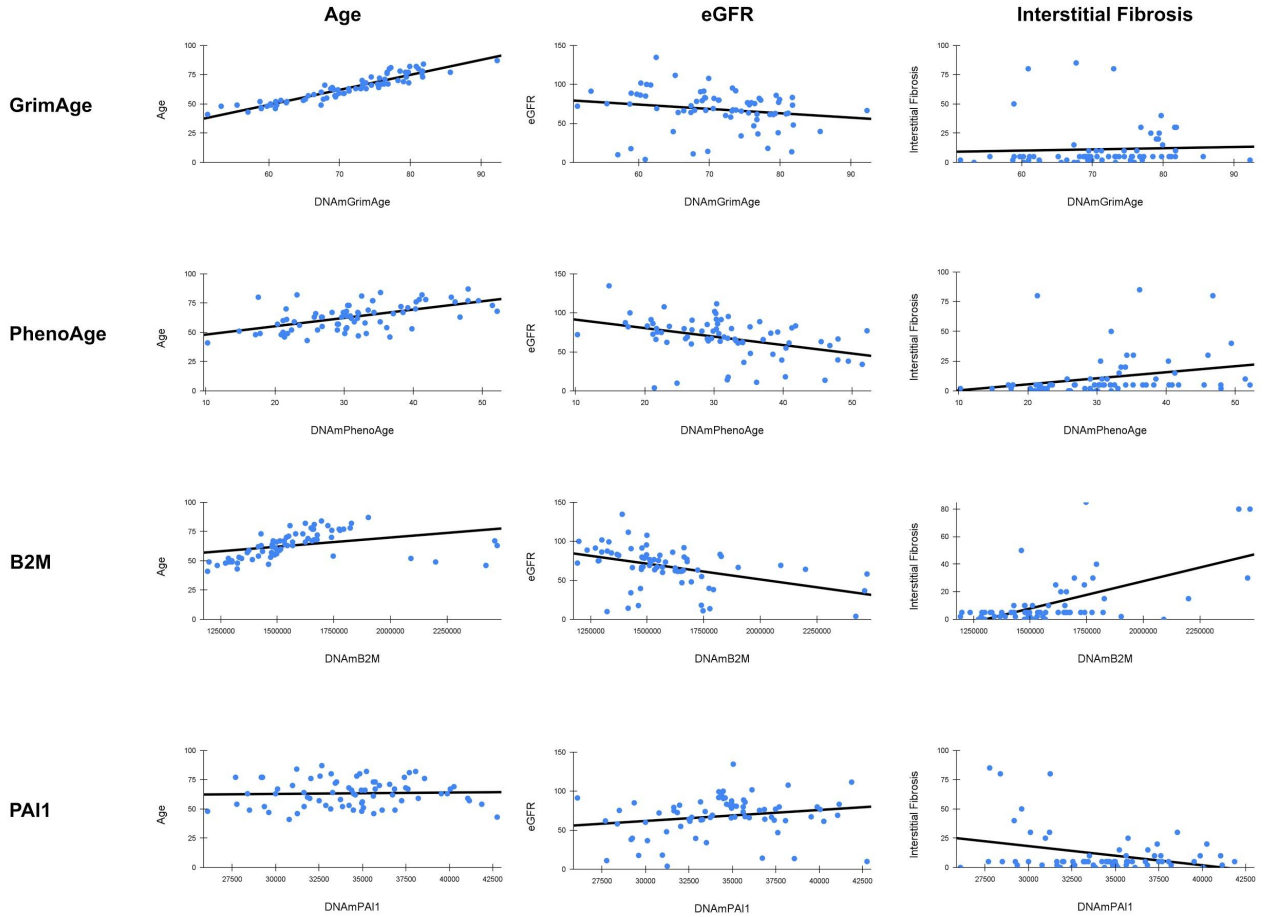
Data availability

Data derived from public domain resources. New analysis data are available in the supplementary material.

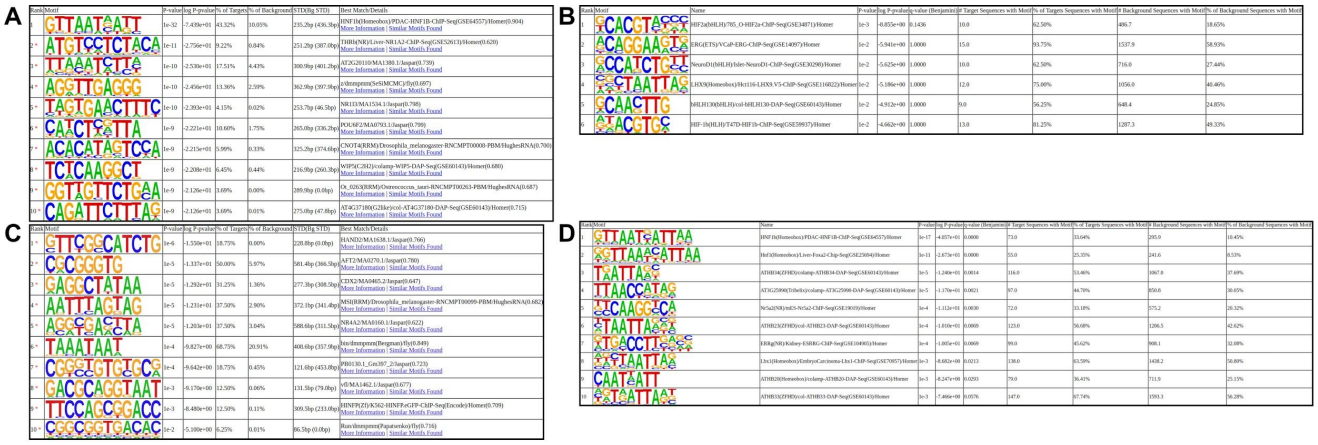
Acknowledgments

We thank the Bar lab members, Dr. Maayan Gal and Prof. Yosef Gruenbaum for critical reading of the manuscript, useful suggestions and comments.

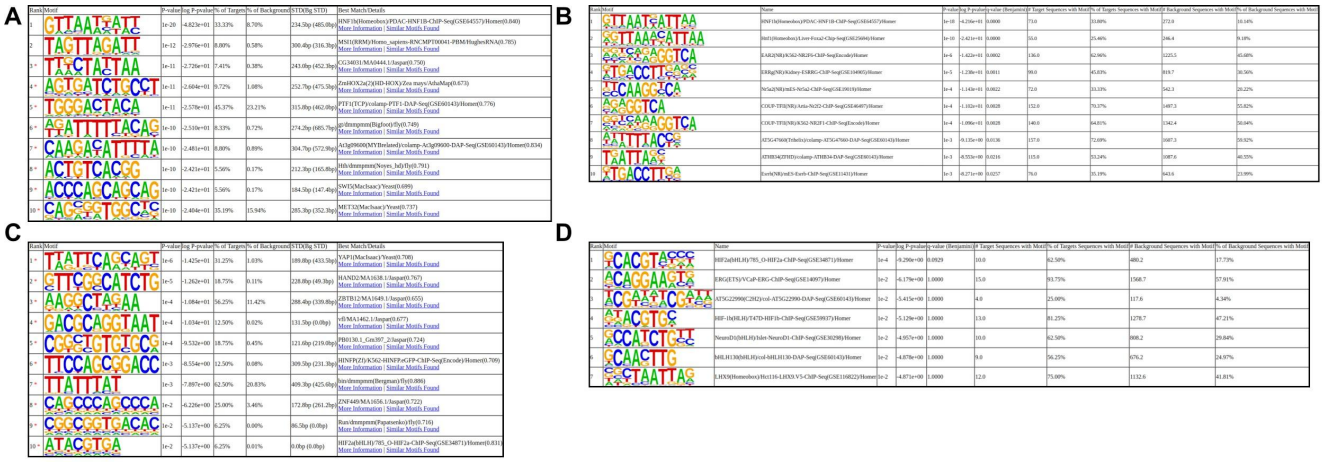
Supplementary figures



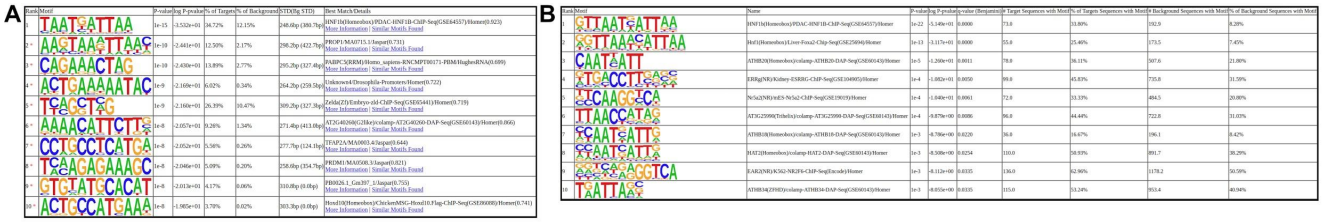
Supplementary figure 1: Methylation clocks correlate with kidney state. Correlation of age, eGFR and IF with four epigenetic clocks: GrimAge, PhenoAge, B2M and PAI1.



Supplementary figure 2: Motif enrichment in uniquely methylated sites vs. random sites. De novo (**A**) and known (**B**) motif enrichment in sites undermethylated in the kidney. De novo (**C**) and known (**D**) motif enrichment in sites overmethylated in the kidney. Up to top 10 results are displayed. * - possible false positive.



Supplementary figure 3: Motif enrichment in uniquely methylated sites vs. IF-correlated sites. De novo (**A**) and known (**B**) motif enrichment in sites undermethylated in the kidney. De novo (**C**) and known (**D**) motif enrichment in sites overmethylated in the kidney. Up to top 10 results are displayed. * - possible false positive.



Supplementary figure 4: Motif enrichment in sites uniquely methylated in kidney vs. other tissues. De novo (**A**) and known (**B**) motif enrichment in sites undermethylated in the kidney. Up to top 10 results are displayed. * - possible false positive.

References

Al-Dabet MM, Shahzad K, Elwakiel A, Sulaj A, Kopf S, Bock F, Gadi I, Zimmermann S, Rana R, Krishnan S, Gupta D, Manoharan J, Fatima S, Nazir S, Schwab C, Baber R, Scholz M, Geffers R, Mertens PR, Nawroth PP, Griffin JH, Keller M, Dockendorff C, Kohli S & Isermann B (2022) Reversal of the renal hyperglycemic memory in diabetic kidney disease by targeting sustained tubular p21 expression. *Nat. Commun.* 13, 5062.

Anon (2013) Chapter 1: Definition and classification of CKD. *Kidney Int. Suppl.* 3, 19–62.

Argyropoulos CP, Chen SS, Ng Y-H, Roumelioti M-E, Shaffi K, Singh PP & Tzamaloukas AH (2017) Rediscovering Beta-2 Microglobulin As a Biomarker across the Spectrum of Kidney Diseases. *Front. Med.* 4, 73.

Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, Slieker RC, Jhamai PM, Verbiest M, Suchiman HED, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Arindrarto W, Kielbasa SM, Jonkers I, van ’t Hof P, Nooren I, Beekman M, Deelen J, van Heemst D, Zhernakova A, Tigchelaar EF, Swertz MA, Hofman A, Uitterlinden AG, Pool R, van Dongen J, Hottenga JJ, Stehouwer CDA, van der Kallen CJH, Schalkwijk CG, van den Berg LH, van Zwet EW, Mei H, Li Y, Lemire M, Hudson TJ, BIOS Consortium, Slagboom PE, Wijmenga C, Veldink JH, van Greevenbroek MMJ, van Duijn CM, Boomsma DI, Isaacs A, Jansen R, van Meurs JBJ, ’t Hoen PAC, Franke L & Heijmans BT (2017) Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* 49, 131–138.

Bontha SV, Maluf DG, Archer KJ, Dumur CI, Dozmorov MG, King AL, Akalin E, Mueller TF, Gallon L & Mas VR (2017) Effects of DNA Methylation on Progression to Interstitial Fibrosis and Tubular Atrophy in Renal Allograft Biopsies: A Multi-Omics Approach. *Am. J. Transplant* 17,

3060–3075.

Brück K, Stel VS, Gambaro G, Hallan S, Völzke H, Ärnlöv J, Kastarinen M, Guessous I, Vinhas J, Stengel B, Brenner H, Chudek J, Romundstad S, Tomson C, Gonzalez AO, Bello AK, Ferrieres J, Palmieri L, Browne G, Capuano V, Van Biesen W, Zoccali C, Gansevoort R, Navis G, Rothenbacher D, Ferraro PM, Nitsch D, Wanner C, Jager KJ & European CKD Burden Consortium (2016) CKD Prevalence Varies across the European General Population. *J. Am. Soc. Nephrol.* 27, 2135–2147.

Christensson A, Savage C, Sjoberg DD, Cronin AM, O'Brien MF, Lowrance W, Nilsson PM, Vickers AJ, Russo P & Lilja H (2013) Association of cancer with moderately impaired renal function at baseline in a large, representative, population-based cohort followed for up to 30 years. *Int. J. Cancer* 133, 1452–1458.

Chu AY, Tin A, Schlosser P, Ko Y-A, Qiu C, Yao C, Joehanes R, Grams ME, Liang L, Gluck CA, Liu C, Coresh J, Hwang S-J, Levy D, Boerwinkle E, Pankow JS, Yang Q, Fornage M, Fox CS, Susztak K & Köttgen A (2017) Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat. Commun.* 8, 1286.

Dania Machlab, Lukas Burger, Charlotte Soneson, Michael Stadler (2021) *monaLisa*, Bioconductor. Available at: https://bioconductor.org/packages/monaLisa.

Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, Coarfa C, Harris RA, Milosavljevic A, Troakes C, Al-Sarraj S, Dobson R, Schalkwyk LC & Mill J (2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol.* 13, R43.

Dhar GA, Saha S, Mitra P & Nag Chaudhuri R (2021) DNA methylation and regulation of gene expression: Guardian of our health. *Nucleus* 64, 259–270.

Dritsoula A, Kislikova M, Oomatia A, Webster AP, Beck S, Ponticos M, Lindsey B, Norman J, Wheeler DC, Oates T & Caplin B (2021) "Epigenome-wide methylation profile of chronic kidney disease-derived arterial DNA uncovers novel pathways in disease-associated cardiovascular pathology." *Epigenetics* 16, 718–728.

El Agha E, Kramann R, Schneider RK, Li X, Seeger W, Humphreys BD & Bellusci S (2017) Mesenchymal Stem Cells in Fibrotic Disease. *Cell Stem Cell* 21, 166–177.

Fairweather J, Findlay M & Isles C (2020) *Clinical Companion in Nephrology*, Springer, Cham.

GBD 2015 Mortality and Causes of Death Collaborators (2016) Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1459–1544.

Giacchino F, Alloatti S, Quarello F, Bosticardo GM, Giraudo G & Piccoli G (1982) The immunological state in chronic renal insufficiency. *Int. J. Artif. Organs* 5, 237–242.

Gluck C, Qiu C, Han SY, Palmer M, Park J, Ko Y-A, Guan Y, Sheng X, Hanson RL, Huang J, Chen Y, Park ASD, Izquierdo MC, Mantzaris I, Verma A, Pullman J, Li H & Susztak K (2019) Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease. *Nat. Commun.* 10, 2461.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J-B, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T & Zhang K (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49, 359–367.

Héberlé É & Bardet AF (2019) Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* 63, 727–741.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass

CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.

Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol.* 14, R115.

Imai S & Kitano H (1998) Heterochromatin islands and their dynamic reorganization: a hypothesis for three distinctive features of cellular aging. *Exp. Gerontol.* 33, 555–570.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyan S & Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41, 178–186.

Kangaspeska S, Stride B, Métivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F & Reid G (2008) Transient cyclical methylation of promoter DNA. *Nature* 452, 112–115.

Ko Y-A, Mohtat D, Suzuki M, Park ASD, Izquierdo MC, Han SY, Kang HM, Si H, Hostetter T, Pullman JM, Fazzari M, Verma A, Zheng D, Greally JM & Susztak K (2013) Cytosine methylation changes in enhancer regions of core pro-fibrotic genes characterize kidney fibrosis development. *Genome Biol.* 14, R108.

Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J & CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) (2009) A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* 150, 604–612.

Levin A & Stevens PE (2011) Early detection of CKD: the benefits, limitations and effects on prognosis. *Nat. Rev. Nephrol.* 7, 446–457.

Lowrance WT, Ordoñez J, Udaltsova N, Russo P & Go AS (2014) CKD and the risk of incident cancer. *J. Am. Soc. Nephrol.* 25, 2327–2334.

Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L & Horvath S (2019) DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 11, 303–327.

Lu Y, Brommer B, Tian X, Krishnan A, Meer M, Wang C, Vera DL, Zeng Q, Yu D, Bonkowski MS, Yang J-H, Zhou S, Hoffmann EM, Karg MM, Schultz MB, Kane AE, Davidsohn N, Korobkina E, Chwalek K, Rajman LA, Church GM, Hochedlinger K, Gladyshev VN, Horvath S, Levine ME, Gregory-Ksander MS, Ksander BR, He Z & Sinclair DA (2020) Reprogramming to recover youthful epigenetic information and restore vision. *Nature* 588, 124–129.

Mancilla E, Avila-Casado C, Uribe-Uribe N, Morales-Buenrostro LE, Rodríguez F, Vilatoba M, Gabilondo B, Aburto S, Rodríguez RM, Magaña S, Magaña F & Alberú J (2008) Time-zero renal biopsy in living kidney transplantation: a valuable opportunity to correlate predonation clinical data with histological abnormalities. *Transplantation* 86, 1684–1688.

Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR & Vinson C (2013) CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Res.* 23, 988–997.

Métivier R, Gallais R, Tiffoche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F & Salbert G (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45–50.

Moore LD, Le T & Fan G (2013) DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23–38.

Musso CG, Álvarez-Gregori J, Jauregui J & Macías-Núñez JF (2016) Glomerular filtration rate equations: a comprehensive review. *Int. Urol. Nephrol.* 48, 1105–1110.

Nakagawa S, Nishihara K, Miyata H, Shinke H, Tomita E, Kajiwara M, Matsubara T, Iehara N, Igarashi Y, Yamada H, Fukatsu A, Yanagita M, Matsubara K & Masuda S (2015) Molecular Markers of Tubulointerstitial Fibrosis and Tubular Cell Damage in Patients with Chronic Kidney Disease. *PLoS One* 10, e0136994.

Okamoto M, Koshino K, Nobori S, Ushigome H, Okajima H, Urasaki K & Yoshimura N (2010) Analysis of preexisting baseline kidney lesions revealed by biopsy in living kidney donors: relationship with clinical parameters at the time of donation. *Clin. Transplant.* 24 Suppl 22, 27–30.

Rakyan VK, Down TA, Balding DJ & Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 12, 529–541.

Rysz J, Franczyk B, Rysz-Górzyńska M & Gluba-Brzózka A (2022) Are Alterations in DNA Methylation Related to CKD Development? *Int. J. Mol. Sci.* 23. Available at: http://dx.doi.org/10.3390/ijms23137108.

Sapienza C, Lee J, Powell J, Erinle O, Yafai F, Reichert J, Siraj ES & Madaio M (2011) DNA methylation profiling identifies epigenetic differences between diabetes patients with ESRD and diabetes patients without nephropathy. *Epigenetics* 6, 20–28.

Schlosser P, Tin A, Matias-Garcia PR, Thio CHL, Joehanes R, Liu H, Weihs A, Yu Z, Hoppmann A, Grundner-Culemann F, Min JL, Adeyemo AA, Agyemang C, Ärnlöv J, Aziz NA, Baccarelli A, Bochud M, Brenner H, Breteler MMB, Carmeli C, Chaker L, Chambers JC, Cole SA, Coresh J, Corre T, Correa A, Cox SR, de Klein N, Delgado GE, Domingo-Relloso A, Eckardt K-U, Ekici AB, Endlich K, Evans KL, Floyd JS, Fornage M, Franke L, Fraszczyk E, Gao X, Gào X, Ghanbari M, Ghasemi S, Gieger C, Greenland P, Grove ML, Harris SE, Hemani G, Henneman P, Herder C, Horvath S, Hou L, Hurme MA, Hwang S-J, Jarvelin M-R, Kardia SLR, Kasela S, Kleber ME, Koenig W, Kooner JS, Kramer H, Kronenberg F, Kühnel B, Lehtimäki T, Lind L, Liu D, Liu Y, Lloyd-Jones DM, Lohman K, Lorkowski S, Lu AT, Marioni RE, März W, McCartney DL, Meeks KAC, Milani L, Mishra PP, Nauck M, Navas-Acien A, Nowak C, Peters A, Prokisch H, Psaty BM, Raitakari OT, Ratliff SM, Reiner AP, Rosas SE, Schöttker B, Schwartz J, Sedaghat S, Smith JA, Sotoodehnia N, Stocker HR, Stringhini S, Sundström J, Swenson BR, Tellez-Plaza M, van Meurs JBJ, van Vliet-Ostaptchouk JV, Venema A, Verweij N, Walker RM, Wielscher M, Winkelmann J, Wolffenbuttel BHR, Zhao W, Zheng Y, Estonian Biobank Research Team, Genetics of DNA Methylation Consortium, Loh M, Snieder H, Levy D, Waldenberger M, Susztak K, Köttgen A & Teumer A (2021) Meta-analyses identify DNA methylation associated with kidney function and damage. *Nat. Commun.* 12, 7174.

Sedighi O, Abediankenari S & Omranifar B (2015) Association between plasma Beta-2 microglobulin level and cardiac performance in patients with chronic kidney disease. *Nephrourol. Mon.* 7, e23563.

Shlipak MG, Fried LF, Crump C, Bleyer AJ, Manolio TA, Tracy RP, Furberg CD & Psaty BM (2003) Elevations of inflammatory and procoagulant biomarkers in elderly persons with renal insufficiency. *Circulation* 107, 87–92.

Silverstein DM (2009) Inflammation in chronic kidney disease: role in the progression of renal and cardiovascular disease. *Pediatr. Nephrol.* 24, 1445–1452.

Smyth LJ, McKay GJ, Maxwell AP & McKnight AJ (2014) DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics* 9, 366–376.

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Müller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T & Vermeulen M (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159.

Steubl D, Block M, Herbst V, Nockher WA, Schlumberger W, Satanovskij R, Angermann S, Hasenau A-L, Stecher L, Heemann U, Renders L & Scherberich J (2016) Plasma Uromodulin

Correlates With Kidney Function and Identifies Early Stages in Chronic Kidney Disease Patients. *Medicine* 95, e3011.

Wing MR, Devaney JM, Joffe MM, Xie D, Feldman HI, Dominic EA, Guzman NJ, Ramezani A, Susztak K, Herman JG, Cope L, Harmon B, Kwabi-Addo B, Gordish-Dressman H, Go AS, He J, Lash JP, Kusek JW, Raj DS & Chronic Renal Insufficiency Cohort (CRIC) Study (2014) DNA methylation profile associated with rapid decline in kidney function: findings from the CRIC study. *Nephrol. Dial. Transplant* 29, 864–872.

Xiong Z, Li M, Ma Y, Li R & Bao Y (2021) GMQN: A Reference-Based Method for Correcting Batch Effects and Probe Bias in HumanMethylation BeadChip. *Front. Genet.* 12, 810985.

X Z, D H, Dg B, Ga H, B B & Pl Z (2014) Urinary β2-microglobulin is a sensitive indicator for renal tubular injury. *SAJ Case Rep.* 1. Available at: http://fulltext.scholarena.com/Urinary-Microglobulin-is-a-Sensitive-Indicator-for-Renal-Tubular-Injury.php.